ZooKeys 921: 141–157 (2020)
doi: 10.3897/zookeys.921.49199
http://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

DNA barcode library for European Gelechiidae (Lepidoptera) suggests greatly underestimated species diversity

Peter Huemer[1], Ole Karsholt[2], Leif Aarvik[3], Kai Berggren[4], Oleksiy Bidzilya[5], Jari Junnilainen[6], Jean-François Landry[7], Marko Mutanen[8], Kari Nupponen[9], Andreas Segerer[10], Jan Šumpich[11], Christian Wieser[12], Benjamin Wiesmair[1], Paul D.N. Hebert[13]

1 *Naturwissenschaftliche Sammlungen, Tiroler Landesmuseen Betriebsges.m.b.H., Innsbruck, Austria* **2** *Zoological Museum, Natural History Museum of Denmark, Copenhagen, Denmark* **3** *Natural History Museum, University of Oslo, Oslo, Norway* **4** *Kristiansand, Norway* **5** *Institute for Evolutionary Ecology of the National Academy of Sciences of Ukraine, Kiev, Ukraine* **6** *Finnish Museum of Natural History, Zoology Unit, Helsinki, Finland* **7** *Canadian National Collection of Insects, Arachnids, and Nematodes, Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada* **8** *Department of Ecology and Genetics, University of Oulu, Finland* **9** *Espoo, Finland* **10** *SNSB-Zoological State Collection, Munich, Germany* **11** *National Museum, Natural History Museum, Department of Entomology, Praha, Czech Republic* **12** *Landesmuseum Kärnten, Klagenfurt, Austria* **13** *Centre for Biodiversity Genomics, University of Guelph, Guelph, Canada*

Corresponding author: Peter Huemer (p.huemer@tiroler-landesmuseen.at)

Academic editor: Mark Metz | Received 8 December 2019 | Accepted 6 February 2020 | Published 24 March 2020

http://zoobank.org/7B93B1C2-D4B0-407D-96FF-E3D5BD4FC415

Citation: Huemer P, Karsholt O, Aarvik L, Berggren K, Bidzilya O, Junnilainen J, Landry J-F, Mutanen M, Nupponen K, Segerer A, Šumpich J, Wieser C, Wiesmair B, Hebert PDN (2020) DNA barcode library for European Gelechiidae (Lepidoptera) suggests greatly underestimated species diversity. ZooKeys 921: 141–157. https://doi.org/10.3897/zookeys.921.49199

Abstract

For the first time, a nearly complete barcode library for European Gelechiidae is provided. DNA barcode sequences (COI gene – cytochrome *c* oxidase 1) from 751 out of 865 nominal species, belonging to 105 genera, were successfully recovered. A total of 741 species represented by specimens with sequences ≥ 500bp and an additional ten species represented by specimens with shorter sequences were used to produce 53 NJ trees. Intraspecific barcode divergence averaged only 0.54% whereas distance to the Nearest-Neighbour species averaged 5.58%. Of these, 710 species possessed unique DNA barcodes, but 31 species could not be reliably discriminated because of barcode sharing or partial barcode overlap. Species

Copyright *Peter Huemer et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

discrimination based on the Barcode Index System (BIN) was successful for 668 out of 723 species which clustered from minimum one to maximum 22 unique BINs. Fifty-five species shared a BIN with up to four species and identification from DNA barcode data is uncertain. Finally, 65 clusters with a unique BIN remained unidentified to species level. These putative taxa, as well as 114 nominal species with more than one BIN, suggest the presence of considerable cryptic diversity, cases which should be examined in future revisionary studies.

Keywords
Europe, cryptic diversity, DNA barcoding, revision, species delimitation

Introduction

The megadiverse family, Gelechiidae, includes approximately 4,700 known species and perhaps a similar number of undescribed taxa (Karsholt et al. 2013). With a remarkable 865 species reported from Europe and adjacent islands (Huemer and Karsholt 2020), the Gelechiidae are the fourth most diverse family of Lepidoptera after the Noctuidae, Geometridae, and Tortricidae in Europe. Due to their general dull-coloured and inconspicuously patterned wings (Fig. 1), and frequently small size, the Gelechiidae have received little attention from lepidopterists, leading to considerable gaps in knowledge of their taxonomy, systematics, biology, and distribution. In particular, the lack of generic revisions in several diverse groups has created the widespread impression of a "difficult" family which has acted to further limit interest in this group.

Over the last two decades, the Gelechiidae have received increasing attention as a result of two monographs that treated approximately half the known European species (Huemer and Karsholt 1999, 2010) and another on the Central European fauna (Elsner et al. 1999). Unfortunately, these publications, as well as several subsequent revisions (i.e., Bidzilya 2005a, 2005b, Bidzilya and Karsholt 2015, Karsholt and Rutten 2005, Karsholt and Šumpich 2015, Li and Sattler 2012), did not take advantage of new molecular methods, in particular DNA barcoding. On the contrary phylogenetic analysis of higher taxa in Gelechiidae benefitted greatly from molecular analysis (Kaila et al. 2011, Karsholt et al. 2013). However, recent studies on several genera of European Gelechiidae (Huemer et al. 2013, 2014, Huemer and Mutanen 2012, Huemer and Karsholt 2014, Landry et al. 2017) revealed the power of this approach to aid species delimitation in taxonomically difficult groups, even those with a high level of unrecorded species and cryptic diversity. Similar patterns have been analyzed in several other Lepidoptera in different parts of the world, e.g., in another gelechioid group (Mutanen et al. 2011), in Iberian butterflies (Dincă et al. 2015), in North American Noctuoidea (Zahiri et al. 2017), or in the Lepidoptera fauna of Costa Rica (Janzen and Hallwachs 2016). These results motivated the present effort to compile a comprehensive DNA barcode library for the European Gelechiidae fauna, with the aim of simplifying future revisionary studies while also improving their quality.

Figure 1. *Megacraspedus teriolensis* is a characteristic example of gelechiid moths only recognised and described during the last few years.

Materials and methods

Checklist of European Gelechiidae

The lack of an updated checklist for European Gelechiidae (see Karsholt 2004-2019) was such a major impediment to the present study that it necessitated the assembly of a new systematic list (Huemer and Karsholt 2020). This list, which includes 865 species of Gelechiidae in 109 genera, provided the basis for selecting the specimens that were analysed in this study.

Sample material

One major challenge was the difficulty in accessing specimens suitable for molecular analysis, reflecting the rarity of many species. In addition, DNA quality of the specimens was another very important limitation as sequence recovery from older specimens of rare taxa was either partial or failed completely even with protocols that employed high-throughput sequencers to analyze short amplicons. In some cases, efforts were made to recollect taxa that lacked a sequence record.

Voucher material was obtained from Europe (Fig. 2) except for eleven taxa whose sequences could not be recovered from specimens from this continent or where it

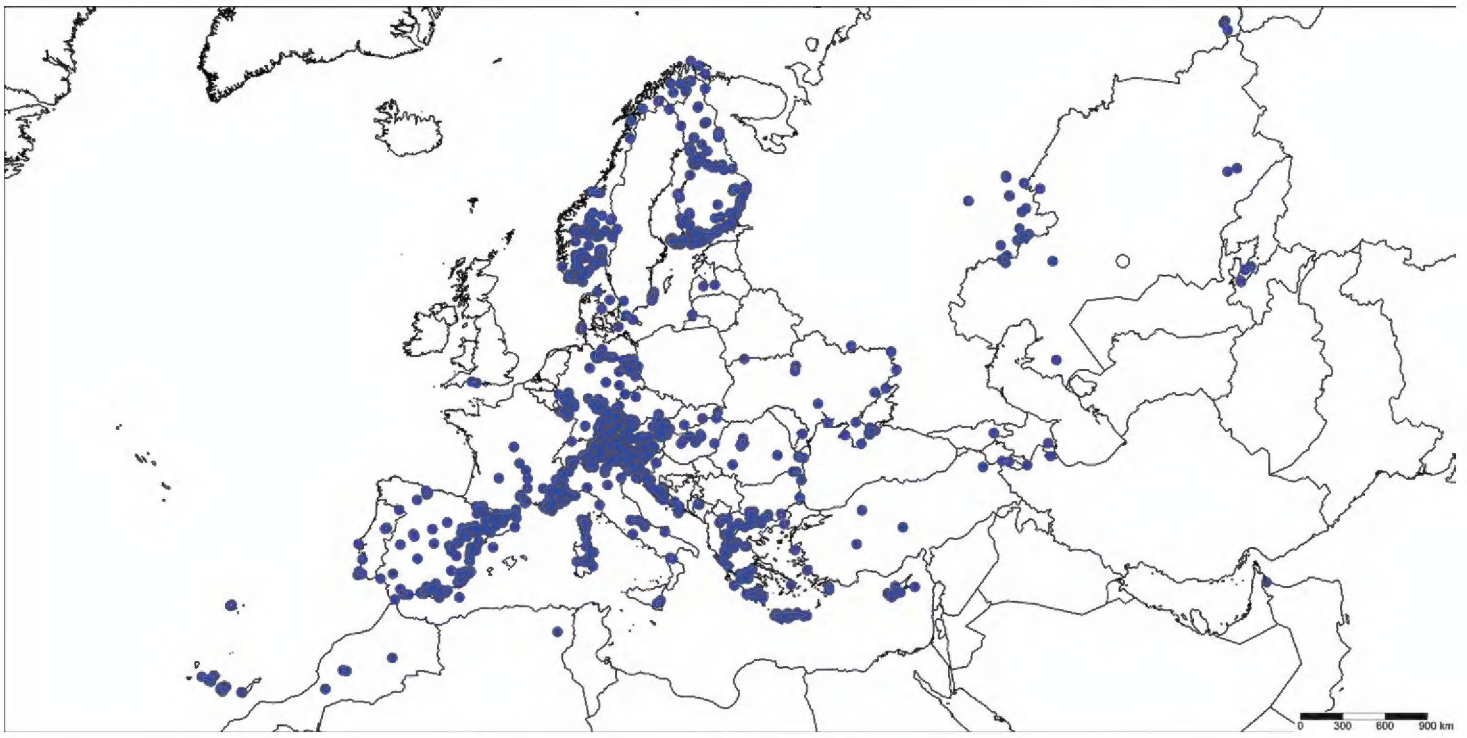

Figure 2. Distribution map of examined material of Gelechiidae (extra-European material partially mapped). SimpleMappr (http://www.simplemappr.net).

seemed important to analyze specimens to clarify taxonomy (e.g., extra-European type-material) (Suppl. material 2, 3). Approximately two-thirds of specimens originated from four nations - Germany (1319), Austria (1157), Italy (906), and Finland (707). The remaining specimens derived from 33 other countries (Fig. 2).

Many institutions and private collectors contributed to the dataset (see below), supplemented by DNA barcodes from earlier studies.

Abbreviations of private and institutional collections

BIOUG Centre for Biodiversity Genomics, Guelph, Canada
INDO Inatura, Dornbirn, Austria
LMK Landesmuseum Kärnten, Klagenfurt, Austria
MFSN Museo Friulano di Storia Natural, Udine, Italy
MZH Finnish Museum of Natural History, Helsinki, Finland
NHM Natural History Museum, London, United Kingdom
NHMO Natural History Museum, University of Oslo, Oslo, Norway
NHMW Naturhistorisches Museum, Vienna, Austria
NMPC National Museum Prague, Czech Republic
NMS Naturmuseum Südtirol, Bozen, Italy
RCAH Research Collection Alfred Haslberger, Teisendorf, Germany
RCER Research Collection Emily Requena Miret, Gurb, Spain
RCGB Research Collection Giorgio Baldizzone, Asti, Italy
RCGT Research Collection Giovanni Timossi, Oderzo, Italy
RCHW Research Collection Hartmut Wegner, Adendorf, Germany
RCIB Research Collection Ian Barton, Cambs, United Kingdom

RCIR Research Collection Ignác Richter, Malá Čausa, Slovakia
RCJD Research Collection Jordi Dantart, Barcelona, Spain
RCJJ Research Collection Jari Junnilainen, Vantaa, Finland
RCJK Research Collection Jari-Pekka Kaitila, Vantaa, Finland
RCJL Research Collection Gérard Labonne, Montpellier, France
RCJN Research Collection Jacques Nel, La Ciotat, France
RCJS Research Collection Jan Skyva, Prague, Czech Republic
RCJSC Research Collection Jürg Schmid, Illanz, Switzerland
RCKB Research Collection Kai Berggren, Kristiansand, Norway
RCKN Research Collection Kari and Timo Nupponen, Espoo, Finland
RCMC Research Collection Martin Corley, Faringdon, U.K.
RCOB Research Collection Oleksiy Bidzilya, Kiev, Ukraine
RCOR Research Collection Oliver Rist, Vienna, Austria
RCPB Research Collection Peter Buchner, Schwarzau am Steinfeld, Austria
RCPL Research Collection Peter Lichtmannecker, Adlkofen, Germany
RCRH Research Collection Robert Heckford, Plympton, Plymouth, U.K.
RCRHE Research Collection Richard Heindel, Günzburg, Germany
RCSP Research Collection Serge Peslier, Perpignan, France
RCTG Research Collection Thomas Guggemoos, Ohlstadt, Germany
RCTM Research Collection Toni Mayr, Feldkirch, Austria
RCTV Research Collection Thierry Varenne, Nice, France
RCWS Research Collection Wolfgang Stark, Trübensee, Austria
RCZT Research Collection Zdenko Tokár, Šaľa, Slovakia
TLMF Tiroler Landesmuseum Ferdinandeum, Innsbruck, Austria
USNM Smithsonian Institution, National Museum of Natural History, Washington DC, U.S.A.
ZMAK Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany
ZMKU Taras Shevchenko National University of Kiev, Kiev, Ukraine
ZMUC Zoological Museum, Natural History Museum of Denmark, Copenhagen, Denmark
ZMUO Zoological Museum, University of Oulu, Finland
ZSM Zoologische Staatssammlung, Munich, Germany

DNA sequencing

A single leg was removed from each specimen and placed in a 96-well lysis plate that was submitted for analysis at the CCDB (Canadian Center for DNA Barcoding, University of Guelph, Canada) where DNA extraction, PCR amplification, and sequencing were performed following standard high-throughput protocols (deWaard et al. 2008). In total, 5986 specimens of European Gelechiidae, initially pre-identified from external and partially genitalia morphology by several colleagues and cross-checked by PH and OK in dubious cases, were successfully sequenced. Details of specimens,

including complete voucher data, images, and GenBank accession numbers are available on BOLD (Ratnasingham 2018, Ratnasingham and Hebert 2007) in the public dataset "Lepidoptera (Gelechiidae) of Europe" under the DOI: dx.doi.org/10.5883/DS-GELECHEU.

Data analysis

Levels of intra- and interspecific variation in the DNA barcode fragment were calculated under the Kimura 2-parameter (K2P) model of nucleotide substitution using analytical tools in BOLD systems v4.0 (http://www.boldsystems.org). Fifty-three Neighbor-Joining trees (Maximum Composite Likelihood method, default settings), most including representatives of a single genus, were constructed using MEGA X (Kumar et. al 2018) (Suppl. material 2 and 3). Node confidences were estimated using 500 bootstrap replicates. For genera with few species, several morphologically closely related genera were included in a single tree. For calculating these trees only sequences ≥ 500 bp were used, except for ten species where only shorter sequences were available (Suppl. material 1). In those cases where the specimens of a single species were assigned to two or more different BINs, they were discriminated by a letter code. Because of the high number of BINs for *Megacraspedus dolosellus* and *M. lanceolellus*, these taxa were figured in two separate NJ trees with BINs separated as single clusters. Species sharing a BIN, but still with a diagnostic barcode were grouped in separate clusters. A three-letter code (ISO 3166-1 alpha-3, https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3) was used to abbreviate country names.

Identification success was assessed by the Barcode Index Number (BIN) system as implemented on BOLD (Ratnasingham and Hebert 2013). This system employs a two-stage algorithm that groups all sequences > 500 bp that meet defined quality criteria into Operational Taxonomic Units (OTUs) and automatically assigns new sequences, irrespective of their previous taxonomy and origin. Concordance or discordance between BINs and morphological species identification was assessed.

Results

Overview

DNA barcode sequences were recovered from 5986 specimens representing 751 of the 865 species of Gelechiidae described from Europe (Suppl. material 1). In addition, the analysis revealed 65 putative species whose members were each assigned to a different unique BIN. Most sequences (5476) were compliant with the barcode standard as described in BOLD (http://www.boldsystems.org). Most subsequent analyses only considered the 741 species with sequences ≥ 500bp, but ten additional species with sequences ≥ 300 bp were included in the NJ trees. Sequences from 723 species qualified for BIN analysis.

Species delimitation from DNA barcode divergences

Intraspecific DNA barcode variation in the 741 named species with sequences ≥500 bp averaged 0.54%, but this may be an underestimate as sample sizes for 224 taxa were low and only represented by singletons. In respect to the distribution of mean intraspecific DNA barcode variation: 73.1% of sequenced species had variation ranging from 0–1%, 15.8% between 1–2%, 6.3% between 2–3%, and 4.8% > 3%.

Contrastingly, barcode gap analysis resulted in mean distances of 5.58% (maximum 12.75%) to the Nearest Neighbor (NN) with only 5.68% of all species showing a NN distance of 0–1% (Table 2). In this latter group, only four species pairs/triplets (*Dirhinosia cervinella* / *D. interposita*, *Iwaruna biguttella* / *I. klimeschi*, *Teleiodes brevivalva* / *T. italica* / *T. vulgella*, *Xenolechia aethiops* / *X. lindae* / *X. pseudovulgella*) shared barcodes so they could not be discriminated on that basis. In eight other cases, shared DNA barcodes meant that assignments were sometimes unreliable, but these species also possessed unique haplotypes (*Acompsia antirrhinella* / *A. tripunctella*, *Anacampsis blattariella* / *A. populella*, *Bryotropha affinis* / *B. umbrosella*, *Sattleria pyrenaica* / *S. melaleucella*, *Scrobipalpa arenbergeri* / *S. mercantourica*, *Stomopteryx lineolella* / *S. nougatricella*, *Thiotricha subocellea* / *T. coleella*, and partially also *Teleiopsis bagriotella* / *T. diffinis* / *T. paulheberti*). Finally, low distances between *Scrobipalpa alterna* / *S. lutea* and *S. halymella* / *S. stabilis* were only based on a single sequence for each of these species so they may represent additional cases of barcode overlap. On the other hand, five other species pairs with low interspecific divergence could be reliably separated by barcodes (*Monochroa arundinetella* / *M. suffusella*, *Scrobipalpa stangei* / *S. artemisiella*, *Scrobipalpa salinella* / *S. salicorniae*, *Scrobipalpula* spp., *Teleiopsis rosalbella* / *T. albifemorella*). Considering all these cases, DNA barcodes showed either incomplete or no resolution for 31 species (4.2%), while species identification was effective for 710 species (95.8%).

Species delimitation with Barcode Index Number (BIN) system

In total, 5877 sequences were assigned to a BIN. These records were assigned to 992 BINs that belong to 788 putative taxa (Suppl. material 2 and 3). Among these, 723 corresponded with named species, while another 65 belong to a unique BIN that is currently unidentified, but many likely represent additional, unrecognised species. Specimens from another 114 named species were assigned to more than one BIN; members of 68 species were placed in two BINs, while BIN counts for the other 46 species ranged from three to 22 (Table 2).

Altogether 668 (92.4%) of 723 named species have one or more unique BINs, while 55 species (7.6%) share a BIN with up to four species (Table 3). BIN sharing was particularly frequent in six genera (*Acompsia*, *Dirhinosia*, *Iwaruna*, *Scrobipalpula*, *Teleiopsis*, *Xenolechia*) where species often cannot be discriminated by DNA barcodes. However, most specimens in these taxa have diagnostic barcodes and all possess diagnostic morphological characters.

Table 1. 42 Species with Nearest-Neighbour distances of 0–1%.

Species	Mean intra-spec.	Max intra-spec.	Nearest species	Dist. NN
Bryotropha affinis	0.17	0.77	*Bryotropha umbrosella*	0
Bryotropha umbrosella	1.76	3.63	*Bryotropha affinis*	0
Iwaruna biguttella	0.78	2.02	*Iwaruna klimeschi*	0
Iwaruna klimeschi	0	0	*Iwaruna biguttella*	0
Teleiodes brevivalva	0.46	0.46	*Teleiodes vulgella*	0
Teleiodes italica	0.32	0.62	*Teleiodes vulgella*	0
Teleiodes vulgella	0.17	0.5	*Teleiodes italica*	0
Xenolechia aethiops	0.08	0.16	*Xenolechia lindae*	0
Xenolechia lindae	0	0	*Xenolechia aethiops*	0
Xenolechia pseudovulgella	N/A	0	*Xenolechia aethiops*	0
Scrobipalpa alterna	N/A	0	*Scrobipalpa lutea*	0.35
Scrobipalpa lutea	N/A	0	*Scrobipalpa alterna*	0.35
Acompsia antirrhinella	1.39	1.39	*Acompsia tripunctella*	0.46
Acompsia tripunctella	2.59	6.4	*Acompsia antirrhinella*	0.46
Dirhinosia cervinella	0.14	0.32	*Dirhinosia interposita*	0.46
Dirhinosia interposita	0	0	*Dirhinosia cervinella*	0.46
Monochroa arundinetella	0.05	0.15	*Monochroa suffusella*	0.47
Monochroa suffusella	0.52	1.07	*Monochroa arundinetella*	0.47
Scrobipalpula psilella	0.21	0.64	*Scrobipalpula seniorum*	0.53
Scrobipalpula seniorum	N/A	0	*Scrobipalpula psilella*	0.53
Anacampsis blattariella	0.48	2.99	*Anacampsis populella*	0.56
Anacampsis populella	0.22	1.41	*Anacampsis blattariella*	0.56
Teleiopsis albifemorella	0.62	1.42	*Teleiopsis rosalbella*	0.61
Teleiopsis bagriotella	0.91	2.66	*Teleiopsis diffinis*	0.61
Teleiopsis diffinis	1.43	3.26	*Teleiopsis bagriotella*	0.61
Teleiopsis rosalbella	0.22	0.46	*Teleiopsis albifemorella*	0.61
Thiotricha coleella	N/A	0	*Thiotricha subocellea*	0.67
Thiotricha subocellea	0.74	1.4	*Thiotricha coleella*	0.67
Stomopteryx lineolella	N/A	0	*Stomopteryx nugatricella*	0.77
Stomopteryx nugatricella	0	0	*Stomopteryx lineolella*	0.77
Scrobipalpula diffluella	0.54	1.2	*Scrobipalpula tussilaginis*	0.8
Scrobipalpula tussilaginis	0.17	0.46	*Scrobipalpula diffluella*	0.8
Scrobipalpa arenbergeri	0.49	0.77	*Scrobipalpa mercantourica*	0.92
Scrobipalpa artemisiella	0.6	2.5	*Scrobipalpa stangei*	0.92
Scrobipalpa mercantourica	N/A	0	*Scrobipalpa arenbergeri*	0.92
Scrobipalpa salicorniae	0.16	0.46	*Scrobipalpa salinella*	0.92
Scrobipalpa salinella	0.28	0.92	*Scrobipalpa salicorniae*	0.92
Scrobipalpa stangei	0.15	0.31	*Scrobipalpa artemisiella*	0.92
Sattleria melaleucella	1.11	1.87	*Sattleria pyrenaica*	0.93
Sattleria pyrenaica	2.6	3.65	*Sattleria melaleucella*	0.93
Scrobipalpa halymella	N/A	0	*Scrobipalpa stabilis*	0.93
Scrobipalpa stabilis	N/A	0	*Scrobipalpa halymella*	0.93

Potential cryptic diversity – unrevised taxa

High levels of 'intraspecific' barcode variation often reflect overlooked species, but there is no fixed level of divergence that indicates species status. Furthermore, deep barcode splits can also arise as a result of the inadvertent recovery of pseudogenes, as a consequence of hybridisation, or *Wolbachia* infection (Mally et al. 2018, Werren et al. 2008). In Lepidoptera, 2–3% divergence is occasionally viewed as signalling the need for further integrative analysis (Hausmann et al. 2013), but there is clear evidence that

Table 2. 46 species of European Gelechiidae assigned to multiple (3-22) BINs

Species	no. of BINs	Species	no. of BINs
Aproaerema anthyllidella	3	*Teleiopsis paulheberti*	3
Aproaerema karvoneni	3	*Aroga flavicomella*	4
Aroga velocella	3	*Caryocolum amaurella*	4
Brachmia dimidiella	3	*Caryocolum fibigerium*	4
Bryotropha desertella	3	*Caryocolum peregrinella*	4
Bryotropha umbrosella	3	*Caryocolum vicinella*	4
Caryocolum alsinella	3	*Ephysteris promptella*	4
Caryocolum marmorea	3	*Gelechia sabinella*	4
Caryocolum tischeriella	3	*Isophrictis anthemidella*	4
Chionodes fumatella	3	*Megacraspedus imparellus*	4
Chionodes viduella	3	*Metzneria metzneriella*	4
Hypatima rhomboidella	3	*Mirificarma cytisella*	4
Isophrictis meridionella	3	*Athrips amoenella*	5
Megacraspedus binotella	3	*Isophrictis kefersteiniellus*	5
Metzneria aprilella	3	*Megacraspedus brachypteris*	5
Metzneria artificella	3	*Monochroa nomadella*	5
Neofaculta ericetella	3	*Sattleria pyrenaica*	5
Oxypteryx baldizzonei	3	*Acompsia tripunctella*	6
Parachronistis albiceps	3	*Caryocolum schleichi*	6
Ptocheuusa paupella	3	*Oxypteryx libertinella*	7
Stomopteryx flavipalpella	3	*Stomopteryx remissella*	8
Teleiodes flavimaculella	3	*Megacraspedus lanceolellus*	20
Teleiodes luculella	3	*Megacraspedus dolosellus*	22

Table 3. Species of European Gelechiidae which share a BIN.

Species	BIN
Acompsia antirrhinella / *A. pyrenaella* / *A. tripunctella*	BOLD:AAJ5937
Anacampsis blattariella / *A. populella*	BOLD:AAD3256
Aproaerema albipalpella / *A. cincticulella*	BOLD:ACB8811
Aristotelia brizella / *A. confusella*	BOLD:AAJ1682
Athrips pruinosella / *A. spiraeae*	BOLD:AAD2577
Caryocolum arenbergeri / *C. blandulella*	BOLD:AAV7765
Dirhinosia cervinella / *D. interposita*	BOLD:ACB0757
Iwaruna biguttella / *I. klimeschi* / *I. robineaui*	BOLD:AAU3602
Metzneria fulva / *M. torosulella*	BOLD:ADM4637
Monochroa arundinetella / *M. suffusella*	BOLD:AAF9390
Monochroa palustrellus / *M. saltenella*	BOLD:AAF2711
Sattleria melaleucella / *S. pyrenaica*	BOLD:AAC5037
Scrobipalpa alterna / *S. lutea*	BOLD:ADR5476
Scrobipalpa amseli / *S. hyssopi*	BOLD:ADL8424
Scrobipalpa artemisiella / *S. stangei*	BOLD:AAE9838
Scrobipalpa halymella / *S. stabilis*	BOLD:AAV9005
Scrobipalpa salicorniae / *S. salinella*	BOLD:AAF1193
Scrobipalpula diffluella / *S. psilella* / *S. ramosella* / *S. seniorum* / *S. tussilaginis*	BOLD:AAF1106
Stomopteryx lineolella / *S. mongolica* / *S. nugatricella*	BOLD:ACB3380
Teleiodes brevivalva / *T. italica* / *T. vulgella*	BOLD:AAE9855
Teleiopsis albifemorella / *T. rosalbella*	BOLD:AAB6930
Teleiopsis bagriotella / *T. diffinis* / *T. paulheberti*	BOLD:ACE4927
Teleiopsis bagriotella / *T. diffinis*	BOLD:ACE6105
Xenolechia aethiops / *X. lindae* / *X. pseudovulgella*	BOLD:AAE1445

no such threshold values exist (see e.g., Kekkonen et al. 2015). In the present dataset 146 of 741 nominal species possessed a maximum intraspecific divergence of > 2%, 88 species > 3%, while 33 species showed greater than > 5% (Table 4).

In some recently revised taxa with high, geographically structured intraspecific barcode divergence such as *Megacraspedus* (Huemer and Karsholt 2018) or the *Oxypteryx libertinella* species-group (Huemer et al. 2013), no evidence for cryptic diversity was found. However, even lower 'intraspecific' barcode divergence may reflect cases of either allopatric or sympatric speciation, as proven e.g., for the genus *Sattleria* (Huemer and Hebert 2011, Huemer and Timossi 2014). In consequence, several species with unusual genetic pattern need to be carefully re-assessed as they may include additional species. Cryptic diversity was, for example, already suspected for some *Caryocolum* (Huemer et al. 2015) or *Stomopteryx remissella*, but may also be detected in recently revised genera such as *Acompsia* or *Chionodes* (Huemer and Karsholt 2002, Huemer and Sattler 1995).

A further group of unrevised species in our dataset includes 65 unidentified DNA barcode clusters which were assigned to separate BINs (Table 5). Many of these cases are likely to represent undescribed species or alternatively, they may represent described species that currently lack barcode coverage. Altogether 26 genera representing approximately one-quarter of European genera are candidates for additional taxa. In fact, four genera (*Aproaerema*, *Aristotelia*, *Monochroa*, *Scrobipalpa*) are each represented by more than five unidentified clusters. For detailed comments on these cases, see Huemer and Karsholt (2020).

Table 4. 33 species of European Gelechiidae with a maximum intraspecific barcode divergence > 5%.

Species	Mean intra-spec.	Max intra-spec.
Megacraspedus dolosellus	7.49	13.76
Megacraspedus lanceolellus	7.37	12.51
Monochroa sepicolella	5.15	9.78
Megacraspedus brachypteris	4.36	7.82
Stomopteryx remissella	2.69	7.47
Ephysteris diminutella	3.87	7.15
Sophronia sicariellus	1.34	7.06
Caryocolum cauligenella	1.86	7.00
Acompsia pyrenaella	3.58	6.92
Caryocolum saginella	2.17	6.86
Dichomeris rasilella	3.31	6.67
Monochroa nomadella	3.72	6.58
Caryocolum schleichi	3.93	6.47
Acompsia tripunctella	2.59	6.40
Megacraspedus teriolensis	3.07	6.38
Caryocolum fibigerium	3.41	6.31
Chionodes fumatella	2.6	6.30
Oxypteryx baldizzonei	3.9	6.29
Oxypteryx wilkella	1.5	6.29
Dichomeris juniperella	2.82	6.24

Species	Mean intra-spec.	Max intra-spec.
Parapodia sinaica	2.97	5.95
Megacraspedus balneariellus	3.97	5.95
Mirificarma burdonella	5.9	5.9
Caryocolum peregrinella	3.56	5.71
Caryocolum alsinella	2.11	5.60
Oxypteryx libertinella	2.65	5.48
Aproaerema suecicella	2.43	5.44
Megacraspedus imparellus	4.05	5.43
Isophrictis anthemidella	2.92	5.3
Catatinagma trivittellum	5.24	5.24
Pexicopia malvella	1.1	5.23
Acompsia maculosella	2.16	5.19
Ephysteris promptella	3.31	5.12

Table 5. Unidentified species of European Gelechiidae with unique BINs.

Taxon	BIN
Anarsia	BOLD:ADE9567
Anarsia	BOLD:ADE9710
Apatetris	BOLD:AAV7596
Apatetris	BOLD:ABA4360
Aproaerema	BOLD:AAT9258
Aproaerema	BOLD:ACF7323
Aproaerema	BOLD:ADG7311
Aproaerema	BOLD:ADL8444
Aproaerema	BOLD:ADL9068
Aproaerema	BOLD:ADL9069
Aristotelia	BOLD:AAU2122
Aristotelia	BOLD:AAV7599
Aristotelia	BOLD:ABV2430
Aristotelia	BOLD:ACC2990
Aristotelia	BOLD:ACK0360
Aristotelia	BOLD:ADC8189
Aristotelia	BOLD:ADK9648
Aristotelia	BOLD:ADL8520
Aristotelia	BOLD:ADL8769
Aristotelia	BOLD:ADL9120
Aristotelia	BOLD:ADM4599
Aristotelia	BOLD:ADY0927
Brachmia	BOLD:ADM5065
Caulastrocecis	BOLD:ADM1812
Caulastrocecis	BOLD:ADR7056
Chrysoesthia	BOLD:ADM8914
Chrysoesthia	BOLD:ADN7772
Dichomeris	BOLD:ADI2574
Epidola	BOLD:ADF2272
Gelechia	BOLD:ADF0061
Gelechiidae	BOLD:ADO2643
Isophrictis	BOLD:ADF3165
Isophrictis	BOLD:ADI3246
Ivanauskiella	BOLD:ACB0708
Megacraspedus	BOLD:ACZ8654
Megacraspedus	BOLD:ADY4582
Mesophleps	BOLD:AAU3614
Mesophleps	BOLD:ADM4492
Metzneria	BOLD:ABW1820
Metzneria	BOLD:ACB3385
Metzneria	BOLD:ADM8252
Monochroa	BOLD:ACF6594
Monochroa	BOLD:ACS5726
Monochroa	BOLD:ACW2532
Monochroa	BOLD:ADL7906
Monochroa	BOLD:ADL9322
Monochroa	BOLD:ADR3927
Neofriseria	BOLD:ADR5460
Ochrodia	BOLD:ACE0260
Oxypteryx	BOLD:ACR9491
Oxypteryx	BOLD:ACS7858
Oxypteryx	BOLD:ACS7859
Psamathocrita	BOLD:ADF0071
Psamathocrita	BOLD:ADL7901
Ptocheuusa	BOLD:AAV7056
Scrobipalpa	BOLD:AAV4547
Scrobipalpa	BOLD:ACT3383
Scrobipalpa	BOLD:ACT4605
Scrobipalpa	BOLD:ADF0070
Scrobipalpa	BOLD:ADG5400
Scrobipalpa	BOLD:ADL6932
Scrobipalpa	BOLD:ADL7117
Sophronia	BOLD:ADF5021
Stomopteryx	BOLD:ADM5270
Telphusa	BOLD:ADM5148

Discussion

During the past decade, several national DNA barcoding campaigns have led to the development of an increasingly well-parameterised DNA barcode library for European Lepidoptera. However, these projects have mainly focused on the fauna of central and northern Europe. As a consequence, genetic coverage for species in the Mediterranean region remains patchy. Reflecting this fact, continent-wide analysis has only considered a few groups so far, such as Nepticulidae (van Nieukerken pers. comm.), Gracillariidae (Lopez-Vaamonde pers. comm.), Elachistinae (Mutanen et al. 2011), Depressariidae (Buchner pers. comm), Geometridae (Hausmann et al. 2013, Müller et al. 2019), and Papilionoidea (Dincă pers. comm.). By contrast, for most families either few DNA barcodes exist, or comprehensive genetic analysis is not available.

The current DNA barcode library makes it clear that the Gelechiidae is a particularly good example of the serious gaps in the knowledge of European biodiversity. Nearly a quarter of current fauna has been described since 1990 (Fig. 3). This gap between European gelechiid diversity and adequate coverage in published alpha-taxonomy is most probably a result of: 1) the small number of gelechiid experts, 2) the lack of adequate vouchers for phenotypic and molecular study 3) the frequently cryptic morphology making them less attractive to non-expert workers, and 4) the infrequent consideration of molecular data to assess taxonomic boundaries.

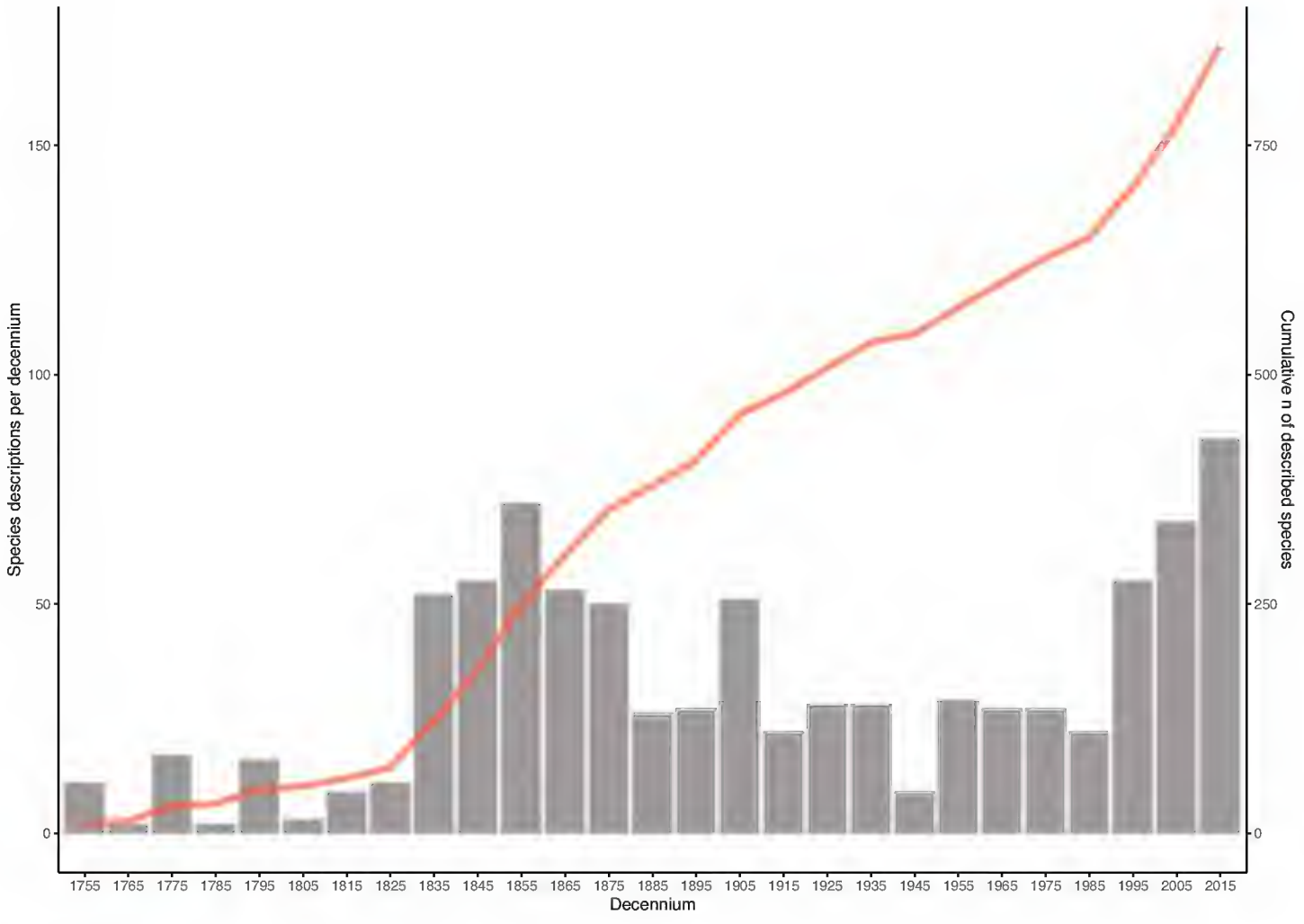

Figure 3. Periods of descriptions of European Gelechiidae.

In the present study, DNA sequences revealed a high level of possible cryptic diversity in European Gelechiidae, despite extensive revisionary work over the last decades (see e.g., Huemer and Karsholt 1999, 2010). Although almost 96% of all 741 species possessed unique barcodes, intraspecific divergences exceeded 2% in nearly a fifth of currently recognised species, and 33 of these cases of divergence values exceeded 5%, values that likely signal overlooked species.

The intraspecific DNA barcode variation is reflected in some taxa as allopatric divergence, but in other cases, it reflects sympatric deep splits. However, few of these species have received detailed taxonomic assessment such as the recent comprehensive study on *Megacraspedus* (Huemer and Karsholt 2018). In many other unrevised genera/species-groups a significant increase in species diversity is likely. The major gaps in taxonomic treatment of European Gelechiidae are further demonstrated by the large number of unidentified genetic clusters revealed by the present investigation as many of these 65 putative taxa are likely to represent undescribed species.

Conclusions

By providing coverage for 751 species of European Gelechiidae, the current DNA barcode library represents the largest release in terms of species diversity for any family of Lepidoptera on this continent. The results reveal unexpected genetic diversity in many taxa as well as numerous unidentified taxa. This indicates that the alpha-taxonomy of this family, still requires serious attention despite one-quarter of the known species described after 1990. The current results indicate that the Gelechiidae remain one of the most taxonomically challenging families of Lepidoptera in the World as complete coverage of even European fauna will require extensive effort. However, the DNA barcode library generated in this study will allow these revisionary studies to target groups that are particularly problematic, accelerating the documentation of the fauna.

Acknowledgements

We are very grateful to the entire team at the Canadian Centre for DNA Barcoding (Guelph, Canada), whose sequencing work was enabled by funding to PDNH from Genome Canada through Ontario Genomics. We also thank the Ontario Ministry of Research and Innovation and NSERC for their support of the BOLD informatics platform. PH is particularly indebted to the Promotion of Educational Policies, University and Research Department of the Autonomous Province of Bolzano - South Tyrol for helping to fund the projects "Genetische Artabgrenzung ausgewählter arktoalpiner und boreomontaner Tiere Südtirols" and "Erstellung einer DNA-Barcode-Bibliothek der Schmetterlinge des zentralen Alpenraumes (Süd-, Nord- und Osttirol)". The work of JS was supported by the Ministry of Culture of the Czech Republic (DKRVO 2019-2023/5.I.a, National Museum, 00023272). MM thanks Finnish Academy, Finnish

Cultural foundation and Kone foundation for financial support to the Finnish Barcode of Life (FinBOL) project.

We gratefully acknowledge the support with material donated to our institutional collections by several colleagues, which was of crucial value for our analysis and to several colleagues mentioned in the list of private and institutional collections) for supporting our DNA barcoding approach.

Philipp Kirschner (Innsbruck, Austria) most kindly helped with the construction of Fig. 3, Romed Unterasinger (TLMF) with the compilation of Suppl. material 2, and Andreas Eckelt (TLMF) with Fig. 1.

This study was only possible due to contributions from many colleagues. As such it provides an impetus for closer co-operation among the community of taxonomists working on Gelechiidae and similarly 'difficult' groups of other micro-moths.

References

Bidzilya O (2005a) A review of the genus *Metanarsia* Staudinger, 1871 (Gelechiidae). Nota lepidopterologica 27: 273–297.

Bidzilya O (2005b) A review of the genus *Athrips* (Lepidoptera, Gelechiidae) in the Palaearctic region. Deutsche Entomologische Zeitschrift 52: 3–72.

Bidzilya O, Karsholt O (2015) Revision of the genus *Istrianis* Meyrick, 1918 (Lepidoptera, Gelechiidae) with special regard to the Palaearctic region. Zootaxa 4059: 401–445. https://doi.org/10.11646/zootaxa.4059.3.1

deWaard JR, Ivanova NV, Hajibabaei M, Hebert PDN (2008) Assembling DNA barcodes: Analytical Protocols. In: Martin C (Ed.) Methods in Molecular Biology:Environmental Genetics. Humana Press, Totowa, NJ, 275–293. https://doi.org/10.1007/978-1-59745-548-0_15

Dincă V, Montagud S, Talavera G, Hernández-Roldán J, Munguira ML, García-Barros E, Hebert PDN, Vila R (2015) DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. Scientific Reports 5: 12395. https://doi.org/10.1038/srep12395

Elsner G, Huemer P, Tokár Z (1999) Gelechiidae Mitteleuropas. Verlag F. Slamka, Bratislava, 208 pp.

Hausmann A, Godfray HCJ, Huemer P, Mutanen M, Rougerie R, van Nieukerken EJ, Ratnasingham S, Hebert PDN (2013) Genetic patterns in European geometrid moths revealed by the Barcode Index Number (BIN) system. PLoS ONE 8(12): e84518. https://doi.org/10.1371/journal.pone.0084518

Huemer P, Elsner G, Karsholt O (2013) Review of the *Eulamprotes wilkella* species group based on morphology and DNA barcodes, with descriptions of new taxa (Lepidoptera, Gelechiidae). Zootaxa 3746: 69–100. https://doi.org/10.11646/zootaxa.3746.1.3

Huemer P, Hebert PDN (2011) Cryptic diversity and phylogeography of high alpine *Sattleria* – a case study combining DNA barcodes and morphology (Lepidoptera: Gelechiidae). Zootaxa 2981: 1–22. https://doi.org/10.11646/zootaxa.2981.1.1

Huemer P, Karsholt O (1999) Gelechiidae I (Gelechiinae: Teleiodini, Gelechiini). In: Huemer P, Karsholt O, Nuss M, Microlepidoptera of Europe. Vol. 3. Apollo Books, Stenstrup, 356 pp.

Huemer P, Karsholt O (2002) A review of the genus *Acompsia* Hübner, 1825, with description of new species (Gelechiidae). Nota lepidopterologica 25(2/3): 109–151.

Huemer P, Karsholt O (2010) Gelechiidae II (Gelechiinae: Gnorimoschemini). In: Huemer P, Karsholt O, Nuss M, Microlepidoptera of Europe. Vol. 6. Apollo Books, Stenstrup, 586 pp. https://doi.org/10.1163/9789004260986

Huemer P, Karsholt O (2018) Revision of the genus *Megacraspedus* Zeller, 1839, a challenging taxonomic tightrope of species delimitation (Lepidoptera, Gelechiidae). ZooKeys 800: 1–278. https://doi.org/10.3897/zookeys.800.26292

Huemer P, Karsholt O (2020) Commented checklist of European Gelechiidae (Lepidoptera). ZooKeys 921: 65–140. https://doi.org/10.3897/zookeys.921.49197

Huemer P, Karsholt O, Mutanen M (2014) DNA barcoding as a screening tool for cryptic diversity: an example from *Caryocolum*, with description of a new species (Lepidoptera, Gelechiidae). ZooKeys 404: 91–111. https://doi.org/10.3897/zookeys.404.7234

Huemer P, Mutanen M (2012) Taxonomy of spatially disjunct alpine *Teleiopsis albifemorella* s. lat. (Lepidoptera: Gelechiidae) revealed by molecular data and morphology – how many species are there? Zootaxa 3580: 1–23. https://doi.org/10.11646/zootaxa.3580.1.1

Huemer P, Sattler K (1995) A taxonomic revision of Palaearctic *Chionodes* (Lepidoptera: Gelechiidae). Beiträge zur Entomologie 45 (1): 3–108.

Huemer P, Timossi G (2014) *Sattleria* revisited: unexpected cryptic diversity on the Balkan Peninsula and in the south-eastern Alps (Lepidoptera: Gelechiidae). Zootaxa 3780: 282–296. https://doi.org/10.11646/zootaxa.3780.2.4

Janzen DH, Hallwachs W (2016) DNA barcoding the Lepidoptera inventory of a large complex tropical conserved wildland, Area de Conservacion Guanacaste, northwestern Costa Rica. Genome 59(9): 641-660. https://doi.org/10.1139/gen-2016-0005

Kaila L, Mutanen M, Nyman T (2011) Phylogeny of the mega-diverse Gelechioidea (Lepidoptera): Adaptations and determinants of success. Molecular Phylogenetics and Evolution 61: 801–809. https://doi.org/10.1016/j.ympev.2011.08.016

Karsholt O (2004–2019) Gelechiidae. In: Karsholt O, van Nieukerken EJ (Eds) Lepidoptera. Fauna Europaea [last update of Gelechiidae: version 2.4 January 2011]. https://fauna-eu.org [Accessed on: 2019-10-14]

Karsholt O, Mutanen M, Lee S, Kaila L (2013) A molecular analysis of the Gelechiidae (Lepidoptera, Gelechioidea) with an interpretative grouping of its taxa. Systematic Entomology 38: 334–348. https://doi.org/10.1111/syen.12006

Karsholt O, Rutten T (2005) The genus *Bryotropha* Heinemann in the western Palaearctic (Lepidoptera: Gelechiidae). Tijdschrift voor Entomologie 148: 77–207. https://doi.org/10.1163/22119434-900000168

Karsholt O, Šumpich J (2015) A review of the genus *Nothris* Hübner, 1825, with description of new species (Lepidoptera: Gelechiidae). Zootaxa 4059: 471–498. https://doi.org/10.11646/zootaxa.4059.3.1

Kekkonen M, Mutanen M, Kaila L, Nieminen M, Hebert PDN (2015) Delineating species with barcodes: a case of taxon dependent method performance in moths. PloS ONE 10(4): e0122481. https://doi.org/10.1371/journal.pone.0122481

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35: 1547–1549. https://doi.org/10.1093/molbev/msy096

Landry J-F, Nazari V, Bidzylia O, Huemer P, Karsholt O (2017) Review of the genus *Agonochaetia* Povolný (Lepidoptera, Gelechiidae), and description of a new genus and species from the Canary Islands. Zootaxa 4300: 451–485. https://doi.org/10.11646/zootaxa.4300.4.1

Li H, Sattler K (2012) A taxonomic revision of the genus *Mesophleps* Hübner, 1825 (Lepidoptera: Gelechiidae). Zootaxa 3373: 1–82. https://doi.org/10.11646/zootaxa.3373.1.1

Mally R, Huemer P, Nuss M (2018) Deep intraspecific DNA barcode splits and hybridisation in the *Udea alpinalis* group (Insecta, Lepidoptera, Crambidae)–an integrative revision. ZooKeys 746: 51–90. https://doi.org/10.3897/zookeys.746.22020

Müller B, Erlacher S, Hausmann A, Rajaei H, Sihvonen P, Skou P (2019) Subfamily Ennominae II. In: Hausmann A, Sihvonen P, Rajaei H, Skou P (2019) The Geometrid Moths of Europe. Vol. 6, Part 1. Brill, Leiden/Boston, 562 pp. https://doi.org/10.1163/9789004387485_001

Mutanen M, Kaila L, Tabell J (2011) Wide-ranging barcoding aids discovery of one-third increase of species richness in presumably well-investigated moths. Scientific Reports 3: 2901. https://doi.org/10.1038/srep02901

Ratnasingham S (2018) BOLD Barcode of Life Data System, version 4. http://www.boldsystems.org [Accessed on: 2019-10-14]

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ratnasingham S, Hebert PN (2013) A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS ONE 8: e66213. https://doi.org/10.1371/journal.pone.0066213

Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. Nature Reviews Microbiology 6: 741–751. https://doi.org/10.1038/nrmicro1969

Zahiri R, Lafontaine JD, Schmidt BC, deWaard JR, Zakharov EV, Hebert PDN (2017) Probing planetary biodiversity with DNA barcodes: The Noctuoidea of North America. PLoS ONE 12(6): e0178548. https://doi.org/10.1371/journal.pone.0178548

Supplementary material 1

Barcoded species

Authors: Peter Huemer, Ole Karsholt, Leif Aarvik, Kai Berggren, Oleksiy Bidzilya, Jari Junnilainen, Jean-François Landry, Marko Mutanen, Kari Nupponen, Andreas Segerer, Jan Šumpich, Christian Wieser, Benjamin Wiesmair, Paul D.N. Hebert

Data type: species data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.921.49199.suppl1

Supplementary material 2

NJ trees 1–53

Authors: Peter Huemer, Ole Karsholt, Leif Aarvik, Kai Berggren, Oleksiy Bidzilya, Jari Junnilainen, Jean-François Landry, Marko Mutanen, Kari Nupponen, Andreas Segerer, Jan Šumpich, Christian Wieser, Benjamin Wiesmair, Paul D.N. Hebert

Data type: species data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.921.49199.suppl2

Supplementary material 3

Specimen data to NJ trees 1–53

Authors: Peter Huemer, Ole Karsholt, Leif Aarvik, Kai Berggren, Oleksiy Bidzilya, Jari Junnilainen, Jean-François Landry, Marko Mutanen, Kari Nupponen, Andreas Segerer, Jan Šumpich, Christian Wieser, Benjamin Wiesmair, Paul D.N. Hebert

Data type: specimen data

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.921.49199.suppl3